

3D 히스토그램 기반 영역분할을 이용한 흉부 X선 영상 품질 평가¹⁾

최현진*, 배수빈**, 박예슬**, 이정원***

*아주대학교 전자공학과

**아주대학교 AI융합네트워크학과

hjaj12020409@ajou.ac.kr, bpgs5316@ajou.ac.kr, yeseuly777@gmail.com,
jungwony@ajou.ac.kr

Quality Evaluation of Chest X-ray Images using Region Segmentation based on 3D Histogram

Hyeon-Jin Choi*, Su-Bin Bea**, Ye-Seul Park**, Jung-Won Lee***

*Dept. of Electrical and Computer Engineering, Ajou University

**Dept. of AI Convergence Network, Ajou University

요 약

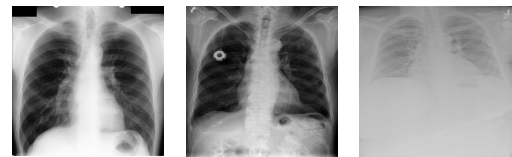
인공지능 기술 발전으로, 의료영상 분야에서도 딥러닝 기반 질병 진단 연구가 활발히 진행되고 있다. 딥러닝 모델 개발 시, 학습 데이터 품질은 모델의 성능과 신뢰성에 매우 큰 영향을 미친다. 그러나 의료 분야의 경우 도메인 지식에 대한 진입 장벽이 높아 개발자가 학습에 사용되는 의료영상 데이터의 품질을 평가하기 어렵다. 이로 인해, 많은 의료영상 분야에서는 각 분야의 특성(질병의 종류, 관찰 아나토미 등)에 따른 영상 품질 평가 방법을 제시해왔다. 그러나 기존의 방법은 특정 질병에 초점이 맞춰져, 일반화된 품질 평가 기준을 제시하고 있지 않다. 따라서 본 논문에서는 대부분의 흉부 질환을 진단하기 위한 흉부 X선 영상의 품질을 평가할 수 있는 기준을 제안한다. 우선, 흉부 X선 영상을 대상으로 관찰된 영역인 심장, 횡격막, 견갑골, 폐 등을 분할하여, 3D 히스토그램을 기반으로 각 영역별 통계적인 정밀 품질 평가 기준을 제안한다. 본 연구에서는 JSRT, Chest 14의 오픈 데이터셋을 활용하여 적용 실험을 수행하였으며, 민감도는 97.6%, 특이도는 92.8%의 우수한 성능을 확인하였다.

1. 서론

인공지능 기술의 발전으로 활용 범위가 광범위해짐에 따라 기계학습 및 딥러닝에 대한 의료 데이터의 적용 및 성능 개선에 대한 연구가 활발히 진행되고 있다. 특히, 의료영상 분야에서도 딥러닝 기술을 통해 질병을 진단하는 연구가 꾸준히 진행되어왔으며, 최근 세계적으로 COVID-19가 확산되어 폐 영상 기반 진단모델 개발이 활발히 이루어지고 있다.

딥러닝 모델 개발 시, 가장 중요시되는 부분은 학습 데이터 품질이다[1]. 딥러닝 모델의 경우, 데이터의 반복적인 예측으로 모델의 손실(Loss)이 결정되고, 손실을 최소화하는 방향으로 모델의 세부 파라미터가 갱신되기 때문이다. 만약 데이터가 심한 노이즈를 갖는다면 모델은 데이터로부터 잘못된 정

보를 학습하게 되고, 잘못된 정보를 학습한 모델은 좋지 않은 성능을 보인다. 이와 같은 특성은 흉부 X선 데이터를 이용한 모델 개발에서도 마찬가지이다. 그림 1은 품질에 따른 흉부 X선 영상을 예시한다.



a. 정상 b. 인공음영 c. 대조도

(그림 1) 품질에 따른 흉부 X선 영상 예시.

흉부 X선 영상에는 의학적으로 위치가 정해진 요소가 있다. 정상 촬영된 흉부 X선 그림 1(a)을 보면 폐, 심장과 횡격막, 견갑골이 두드러지는데, 이들 세 가지는 모든 사람에게서 크기와 간격의 차이는 있으나 같은 배열로 보이는 요소이다. 그림 1(b), (c)는 인공음영과 대조도 측면에서 품질이 낮은 사례이다. X선 투과를 방해하는 인공 장치가 부착 또는 삽입된 경우, 인공음영이 발생하게 된다. 그림 1(b)의 좌측 폐에서 의료 장치 부착에 의한 인공음영을 확

1) “본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음” (IITP-2021-2020-0-01461)

인할 수 있으며, 이러한 노이즈는 질병으로 오인될 수 있다. 또한, 그림 1(c)에서 대조도가 적절하지 않아 폐가 선명하지 않은 예시를 볼 수 있다. 여기서 대조도는 가장 어두운 부분과 가장 밝은 부분의 밝기 비율을 의미한다. 이러한 데이터는 임상에서도 진단에 활용되기 어려우며 딥러닝 모델 개발 시 오히려 잘못된 정보를 학습시킬 가능성이 높다.

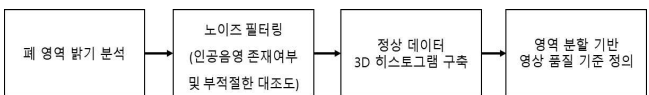
딥러닝 모델이 적용되는 다른 분야와 달리, 의료 분야의 경우 도메인 지식에 대한 진입 장벽이 높아 개발자가 학습에 사용되는 데이터의 품질을 결정하기 어렵다. 이로 인해, 많은 의료영상 분야에서는 각 분야의 특성(질병의 종류, 관찰 아나토미 등)을 반영한 품질 평가 방법을 제시해왔다. [2]는 COVID-19 가능성을 예측하는 CNN의 성능 향상을 위하여 횡경막 영역 제거 전처리 방법을 제안하며, [3]은 질병 분류를 위해 각 질병이 정확히 보이도록 각각에 맞는 전처리 방법을 제안한다. [4]는 폐렴을 식별하기 위해, [5]는 COVID-19 탐지를 위해 폐부분만 남도록 자르는 전처리를 제안한다.

이렇듯 현재 알려진 데이터 품질 관련 연구는 특정 질병 진단, 분류 등 모델 개발을 위한 전처리로 각 질병 모델에 알맞게 개발되어왔다. 그러나 특정 질병이나 관찰 영역에만 초점을 맞추어, 일반화된 품질 평가 기준을 제시하고 있지 않다. 따라서 본 논문에서는 흉부 X선 영상의 관찰 영역을 분할하여, 3D 히스토그램을 기반으로 영역별 통계적인 정밀 품질 평가 기준을 제안한다.

오픈 데이터셋 JSRT[6], Chest 14[7]를 활용, 적용 실험을 수행하였으며, 민감도(Sensitivity) 97.6%, 특이도(Specificity) 92.8%의 우수한 성능을 확인하였다. 제안하는 품질 평가 방법은 모델 설계에 앞서 학습 데이터의 품질을 평가하고, 저품질의 영상을 걸러내어 진단모델의 성능 향상을 지원할 수 있다.

2. 흉부 X선 영상 품질 평가 방법

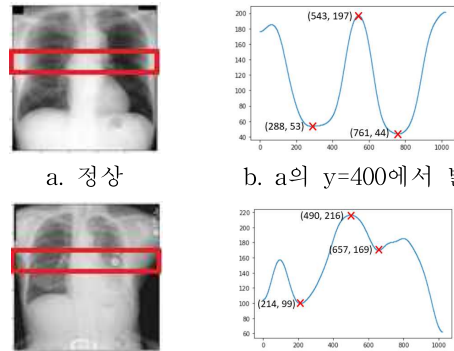
품질 평가 데이터로 활용될 수 있는 고품질의 데이터셋을 확보하기 위해 폐 경계 밝기 변화 분석을 수행하여 인공음영, 대조도의 측면에서 저품질인 데이터를 걸러낸다. 다음은 정제된 데이터에 대해, 영역분할을 통한 정밀 품질 평가 기준을 확립한다. 이를 요약하면 다음 그림 2와 같다.



(그림 2) 흉부 X선 영상 품질 평가 기준 정의 절차

2.1 폐 경계 밝기 변화 분석을 통한 노이즈 검출

고품질의 흉부 X선 영상 기준을 정밀 분석하기 위해, 앞서 설명한 낮은 품질의 노이즈 영상(예: 인공음영 및 대조도) 검출 과정을 선행하여 정상 데이터를 걸러낸다. 폐 영역은 대조도에 따라 갈비뼈나 혈관이 인공음영과 비슷한 밝기로 보일 수 있다. 따라서 폐 경계의 밝기 변화를 분석하여, 인공음영 이미지를 검출한다. 먼저, 흉부 X선 영상에서 좌측 맨 위 지점의 좌표를 $(x, y)=(0, 0)$ 으로, 우측 맨 하단의 좌표를 $(1023, 1023)$ 으로 설정한다. 폐 영역에서 쇄골과 심장은 인공음영과 같이 밝게 보여 밝기 변화 분석에 방해가 되므로, 이를 제외한 영역($x: 0-1023, y: 250-400$)에서 폐 경계 밝기 변화를 분석한다. 그 중 $y=400$ 일 때를 예로 정상, 노이즈 영상에 대해 그래프를 그려본 결과는 그림 3(b), (d)와 같다.



c. 노이즈(인공음영, 대조도) d. c의 $y=400$ 에서 밝기 분포
(그림 3) 폐 경계 밝기 변화

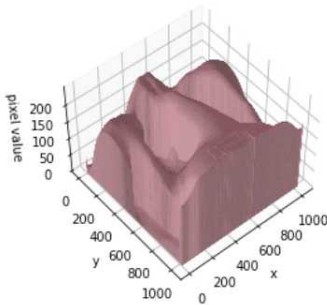
그림 3(b), (d)는 폐 경계 밝기 변화 그래프로, X축은 흉부 X선 영상의 x값(0-1023), Y축은 흉부 X선 영상의 좌표 $(x, 400)$ 에서의 화소값(0-255)이다. 극댓값의 위치와 개수로 인공음영을 판별할 수 있고 극댓값과 극솟값의 차의 크기로 대조도를 판별할 수 있다. 그 예로 정상인 그림 3(b) 폐 극솟값(288, 53), (761, 44)와 척추 극댓값(543, 197)의 차는 144, 153인데 반해 대조도가 낮은 그림 3(d)에서 차는 각각 117, 47로 상대적으로 낮은 값을 보인다. 이를 통해, 흉부 X선 데이터셋에서 인공음영의 유무, 대조 정도를 판별하고, 고품질의 데이터를 정제하여 품질 평가 기준을 확보한다.

2.2 흉부 영역분할 기반 품질 평가 기준 정립

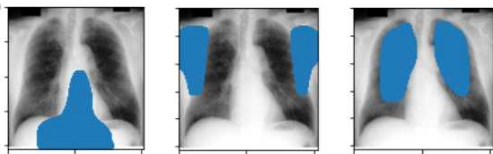
그림 3은 고품질 흉부 X선 영상 데이터셋의 각 이미지를 폐만 드러나도록 자른 뒤, 사이즈를 동일하게 재조정하여 위치별 평균을 구한 결과를 3D 히스토그램으로 표현한 것이다. 영상을 회색조

(gray-scale)로 정규화하였기 때문에 모든 화소는 0에서 255 사이 값을 갖는다. 여기서, 모든 사람에 대해 같은 배열로 보이는 폐, 심장과 횡경막, 견갑골의 영역은 그림 4의 밝기 분포에 따라 구분될 수 있다.

따라서 본 논문에서는 평균 밝기 3D 히스토그램 데이터를 이용하여, 영역을 분할한다. 영역분할을 위해, 그림 4의 Z축(평균 밝기) 데이터를 군집화하였으며, Z축 높이에 따라 같은 군집으로 묶인 영역은 그림 5와 같이 구분된다. 밝기가 큰 순서로 살펴보면, 그림 5(a) 심장과 횡경막, 그림 5(b) 견갑골, 그림 5(c) 폐로 분류된다. 이렇게 정의된 세 영역을 기반으로 고품질 평균 밝기 값을 품질 기준으로 정의, 통합하여 흉부 X선 영상의 품질을 평가한다.



(그림 4) 고품질 흉부 X선 영상의 평균 밝기 분포



a. 심장, 횡경막 b. 견갑골 c. 폐

(그림 5) 흉부 X선 위치별 평균 밝기에 따른 영역분할

3. 실험 결과

흉부 X선 오픈 데이터셋을 활용하여 제안하는 영상 품질 평가 방법의 효용성을 검증한다.

3.1 오픈 데이터셋 사양 분석

실험에서 사용한 흉부 X선 오픈 데이터셋은 JSRT와 Chest14이다. JSRT 데이터셋은 총 247개로 소량이지만, 공신력있는 일본 방사선 학회에서 제공한 양질의 데이터로서 대부분 인공음영이 없고 대조도가 적절하며 환자의 자세도 모두 정면이다. 반면, Chest14 데이터셋은 112,120개로 다량이지만, 인공음영을 포함하며 대조도가 적절하지 않거나 자세가 정면이 아닌 이미지가 다수 포함되어있다. 따라서 본 연구에서는 JSRT 데이터셋 자체를 정상 촬영된 고품질 데이터셋으로 정의하였다. Chest 14 데이터셋은 매우 다량이므로 그중 일부를 사용하였으며 인공

음영, 대조도, 환자의 자세 측면에서 명확히 노이즈 영상으로 판단되는 데이터 2000개를 대상으로 다음 그림과 같이 인공음영, 대조도, 환자의 자세로 노이즈 영상을 레이블링 하였다.

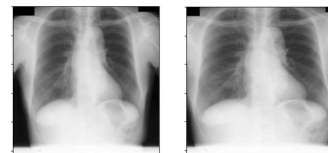
<표 1> 데이터 레이블

분류	데이터셋	개수	
정상 데이터	JSRT	247	
노이즈 데이터	Chest 14 (일부)	2000	
		인공음영	1396
		대조도	540
		환자의 자세	64

3.2 흉부 X선 영상 전처리 및 정규화

품질 평가 실험에 앞서 수집된 데이터셋에 대한 전처리와 정규화 작업을 수행하였다. JSRT의 크기(shape)는 (1024, 1024, 3), Chest14의 크기(shape)는 (2048, 2048, 3)로 두 데이터셋 모두 RGB 색 공간에서 표현된다. 그러나 흉부 X선 이미지는 색조가 다양하지 않으므로, 두 데이터셋의 크기를 회색조로 변형하여, (1024, 1024)로 재설정하였다.

흉부 X선 영상은 환자에 따라 몸과 배경 비율이 다르다. 이는 정상 데이터의 위치별 평균 밝기 분석에 방해요소로 작용하므로 척추 기준 좌우로 가장 밝은 지점을 몸의 윤곽으로 설정, 크기를 조정한다. 폐의 경계는 하단으로 갈수록 좌우로 멀어지므로 폐 영역의 최대한 하단(y=540)을 기준으로 잡는다. 그림 6은 전처리 및 정규화한 흉부 X선 영상 예이다.



a. 전처리 및 정규화 전 b. 전처리 및 정규화 후

(그림 6) 흉부 X선 영상 전처리 및 정규화 예시

3.3 노이즈 영상 검출 결과

다음은 전처리와 정규화를 거친 데이터셋에 대해 폐 경계 밝기 변화 분석을 진행한다. 폐 경계 영역 검출을 위해, 정규화된 전체 위치(1024) 대비 흉곽 위치를 250에서 400 범위에 대해 10단위로 설정하였다(y' = [250, 260, ... , 400]). 정상 데이터를 분석한 결과, 밝기 변화 그래프에서 극댓값은 최소 세 개(척추, 좌우 윤곽)에서 최대 다섯 개(이미지 자체 테두리 번짐)이다. 또한, 극솟값은 척추를 기준으로 좌우 최소 하나 이상 존재한다. 척추 위치 극댓값은 250<=x<=650 범위에 존재하며 이 값과 폐위치 극솟값의 차는 통계적으로 측정된 값인 60 이상이다.

3.4 흉부 영역분할 기반 품질 평가 결과

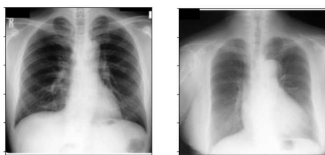
제안한 방법을 기반으로 흉부를 총 3개의 영역으로 분할하고, 정상 데이터에 대한 통계분석을 통해 영역별 품질 평가 기준을 확립하였다. 분할된 영역에서 평가 기준치 이상, 또는 이하인 화소 개수가 오차범위 이상인 데이터를 노이즈로 분류한다. 영역별 품질 평가 기준은 표 2와 같으며 제안하는 품질 평가 방법의 민감도는 97.6%, 특이도는 92.8%이다.

<표 2> 분할된 영역별 품질 평가 임계치

	영역 설정 조건	평가기준	오차범위
심장, 횡경막	$225 \leq m < 256$	$n < 8$	5%
견갑골	$173 \leq m < 225$ $(x < 300) \text{ or } (x > 730)$ $100 < y < 600$	$n < 25$	7%
폐	$50 \leq m < 110$ $190 < x < 870$ $90 < y < 700$	$n > 240$	25%

* m은 평균 밝기값, n은 화소값.

그림 7은 분류된 결과의 사례이다. 정상 데이터 평가 결과 정상으로 검출된 그림 7(a)는 환자의 자세, 인공음영 대조도 모두 노이즈 기준에 해당하지 않아 정상으로 평가되었다. 정상 데이터로 분류하였으나 노이즈로 검출된 그림 7(b)는 노이즈 평가 기준에 해당하지 않지만, 밝기 분석 범위(y: 250-400)에 포함된 쇄골이 인공음영으로 인식되어 노이즈로 평가되었다. 이 예시는 흉곽이 일반적인 데이터와 다르게 위치할 경우 잘못 평가될 수 있음을 시사한다. 노이즈 데이터로 분류하였으나 정상으로 검출된 그림 7(c)는 의료 장치에 의한 인공음영을 갖지만, 그 밝기가 낮아 밝기 분석과정에서 정상 데이터로 평가되었다. 이 예시로 인공음영의 밝기가 낮은 경우 잘못 평가될 수 있음을 보인다. 그림 7(d)는 선명한 인공음영이 있어, 노이즈로 평가되었다.



a. TP(True Positive) b. FN(False Negative)



c. FP(False Positive) d. TN(True Negative)

(그림 7) 정상, 노이즈 데이터 검출 예시

사례 분석 결과 폐가 밝기 분석 범위에 위치하지 않을 경우 FN이 발생하며, 인공음영의 밝기가 어두울 경우 FP가 발생함을 확인할 수 있었다.

4. 결론

본 논문에서는 흉부 X선 영상의 관찰 영역을 분할하여 3D 히스토그램을 기반으로 영역별 통계적인 정밀 평가 기준을 제안한다. 오픈 데이터셋 JSRT, Chest14를 활용하여 적용 실험을 수행한 결과 민감도는 97.6%, 특이도는 92.8%로 우수한 성능을 확인하였다. 향후, 여백에 대한 전처리와 정밀한 밝기 변화 분석을 진행하여 임계치에 대한 일반화된 정의를 개발할 것이다.

참고문헌

[1] Richard Y. Wang, Diane M.Strong, "Beyond Accuracy: What Data Quality Means to Data Consumers", Journal of Management Information Systems , Volume 12, 2015

[2] Morteza Heidari, Seyedehnafiseh Mirniaharikandehei, Abolfazl Zargari Khuzani, Gopichandh Danala, Yuchen Qiu, Bin Zheng, "Improving the performance of CNN to predict the likelihood of COVID-19 using chest X-ray images with preprocessing algorithms", International Journal of Medical Informatics, Volume 144, 2020

[3] Guy Caseneuve, Iren Valova, Nathan LeBlanc, Melanie Thibodeau, "Chest X-Ray Image Preprocessing for Disease Classification", Procedia Computer Science, Volume 192, 658-665, 2021

[4] Khatri A., Jain R., Vashista H., Mittal N., Ranjan P., Janardhanan R. "Pneumonia Identification in Chest X-Ray Images Using EMD", vol 99. Springer, Singapore, 2020

[5] S. Rajaraman, J. Siegelman, P. O. Alderson, L. S. Folio, L. R. Folio and S. K. Antani, "Iteratively Pruned Deep Learning Ensembles for COVID-19 Detection in Chest X-Rays," in IEEE Access, vol. 8, pp. 115041-115050, 2020

[6] JSRT (Japanese Society of Radiological Technology), <http://db.jsrt.or.jp/eng.php>

[7] Chest14 (NIH Chest X-ray Dataset of 14 Common Thorax Disease Categories), <https://www.kaggle.com/nih-chest-xrays/data>